

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: LOWE, JOHN B.
- (ii) TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS, GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
- (iii) NUMBER OF SEQUENCES: 14
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
 - (B) STREET: 1755 Jefferson Davis Highway, Fourth Floor
 - (C) CITY: Arlington
 - (D) STATE: Virginia
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 22202
- (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER: US
 - (B) FILING DATE: 20-JUL-1992
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
- (A) NAME: Lavalleye, Jean-Paul M. P.
 - (B) REGISTRATION NUMBER: 31,451
 - (C) REFERENCE/DOCKET NUMBER: 2363-060-55
- (ix) TELECOMMUNICATION INFORMATION:
- (A) TELEPHONE: (703)521-4500
 - (B) TELEFAX: (703)486-2347
 - (C) TELEX: 248855 OPAT UR

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2043 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AGGAAACCTG CCATGGCCTC CTGGTGAGCT GTCCTCATCC ACTGCTCGCT GCCTCTCCAG 60
ATACTCTGAC CCATGGATCC CTTGGGTGCA GCCAAGCCAC AATGGCCATG GCGCCGCTGT 120
CTGGCCGCAC TGCTATTTCA GCTGCTGGTG GCTGTGTGTT TCTTCTCCTA CCTGCGTGTG 180
TCCCAGACG ATGCCACTGG ATCCCCTAGG GCTCCCAGTG GGTCCCTCCCG ACAGGACACC 240
ACTCCCACCC GCCCCACCCT CTTGATCCTG CTATGGACAT GGCCTTTCCA CATCCCTGTG 300
GCTCTGTCCC GCTGTTCAGA GATGGTGCCC GGCACAGCCG ACTGCCACAT CACTGCCGAC 360
CGCAAGGTGT ACCCACAGGC AGACACGGTC ATCGTGCACC ACTGGGATAT CATGTCCAAC 420
CCTAAGTCAC GCCTCCCACC TTCCCCGAGG CCGCAGGGGC AGCGCTGGAT CTGGTTCAAC 480
TTGGAGCCAC CCCCTAACTG ~~CCAGCACCTG~~ GAAGCCCTGG ~~ACAGATACTT~~ CAATCTCACC 540
ATGTCCTACC GCAGCGACTC ~~CGACATCTTC~~ ACGCCCTACG GCTGGCTGGA GCCGTGGTCC 600
GGCCAGCCTG CCCACCCACC ~~GCTCAACCTC~~ TCGGCCAAGA CCGAGCTGGT GGCCTGGGCG 660
GTGTCCAAC GGAAGCCGGA CTCAGCCAGG GTGCGCTACT ACCAGAGCCT GCAGGCTCAT 720
CTCAAGGTGG ACGTGTACGG ACGCTCCCAC AAGCCCCTGC CCAAGGGGAC CATGATGGAG 780
ACGCTGTCCC GGTACAAGTT CTACCTGGCC TTCGAGAACT CCTTGCACCC CGACTACATC 840
ACCGAGAAGC TGTGGAGGAA CGCCCTGGAG GCCTGGGCCG TGCCCGTGGT GCTGGGCCCC 900
AGCAGAAGCA ACTACGAGAG GTTCCTGCCA CCCGACGCCT TCATCCACGT GGACGACTTC 960
CAGAGCCCCA AGGACCTGGC CCGGTACCTG CAGGAGCTGG ACAAGGACCA CGCCCGCTAC 1020
CTGAGCTACT TTCGCTGGCG GGAGACGCTG CGGCCTCGCT CCTTCAGCTG GGCCTGGAT 1080
TTCTGCAAGG CCTGCTGGAA ACTGCAGCAG GAATCCAGGT ACCAGACGGT GCGCAGCATA 1140
GCGGCTTGGT TCACCTGAGA GGCCGGCATG GTGCCTGGGC TGCCGGGAAC CTCATCTGCC 1200
TGGGGCCTCA CCTGCTGGAG TCCTTTGTGG CCAACCCTCT CTCTTACCTG GGACCTCACA 1260
CGCTGGGCTT CACGGCTGCC AGGAGCCTCT CCCCTCCAGA AGACTTGCCT GCTAGGGACC 1320

TCGCCTGCTG GGGACCTCGC CTGTTGGGGA CCTCACCTGC TGGGGACCTC ACCTGCTGGG 1380
GACCTTGGCT GCTGGAGGCT GCACCTACTG AGGATGTCGG CGGTCGGGGA CTTTACCTGC 1440
TGGGACCTGC TCCCAGAGAC CTTGCCACAC TGAATCTCAC CTGCTGGGGA CCTCACCCTG 1500
GAGGGCCCTG GGCCCTGGGG AACTGGCTTA CTTGGGGCCC CACCCGGGAG TGATGGTTCT 1560
GGCTGATTTG TTTGTGATGT TGTTAGCCGC CTGTGAGGGG TGCAGAGAGA TCATCACGGC 1620
ACGGTTTCCA GATGTAATAC TGCAAGGAAA AATGATGACG TGTCTCCTCA CTCTAGAGGG 1680
GTTGGTCCCA TGGGTTAAGA GCTCACCCCA GGTTCTCACC TCAGGGGTTA AGAGCTCAGA 1740
GTTCAGACAG GTCCAAGTTC AAGCCCAGGA CCACCACTTA TAGGGTACAG GTGGGATCGA 1800
CTGTAAATGA GGAATTCTGG AACATTCCAA ATATTCTGGG GTTGAGGGAA ATTGCTGCTG 1860
TCTACAAAAT GCCAAGGGTG GACAGGCGCT GTGGCTCACG CCTGTAATTC CAGCACTTTG 1920
GGAGGCTGAG GTAGGAGGAT TGATTGAGGC CAAGAGTTAA AGACCAGCCT GGTCAATATA 1980
GCAAGACCAC GTCTCTAAAT AAAAAATAAT AGGCCGGCCA GGAAAAAAAA AAAAAAAAAA 2040

AAA

204

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Asp	Pro	Leu	Gly	Ala	Ala	Lys	Pro	Gln	Trp	Pro	Trp	Arg	Arg	Cys
1				5					10				15		
Leu	Ala	Ala	Leu	Leu	Phe	Gln	Leu	Leu	Val	Ala	Val	Cys	Phe	Phe	Ser
			20					25					30		
Tyr	Leu	Arg	Val	Ser	Arg	Asp	Asp	Ala	Thr	Gly	Ser	Pro	Arg	Ala	Pro
		35					40					45			
Ser	Gly	Ser	Ser	Arg	Gln	Asp	Thr	Thr	Pro	Thr	Arg	Pro	Thr	L u	Leu
	50					55					60				

Ile 65	Leu	Leu	Trp	Thr	Trp 70	Pro	Phe	His	Ile	Pro 75	Val	Ala	Leu	Ser	Arg 80
Cys	Ser	Glu	Met	Val 85	Pro	Gly	Thr	Ala	Asp 90	Cys	His	Ile	Thr	Ala	Asp 95
Arg	Lys	Val	Tyr 100	Pro	Gln	Ala	Asp	Thr 105	Val	Ile	Val	His	His 110	Trp	Asp
Ile	Met	Ser 115	Asn	Pro	Lys	Ser	Arg 120	Leu	Pro	Pro	Ser	Pro 125	Arg	Pro	Gln
Gly	Gln 130	Arg	Trp	Ile	Trp	Phe 135	Asn	Leu	Glu	Pro	Pro 140	Pro	Asn	Cys	Gln
His 145	Leu	Glu	Ala	Leu	Asp 150	Arg	Tyr	Phe	Asn	Leu 155	Thr	Met	Ser	Tyr	Arg 160
Ser	Asp	Ser	Asp	Ile 165	Phe	Thr	Pro	Tyr	Gly 170	Trp	Leu	Glu	Pro	Trp 175	Ser
Gly	Gln	Pro	Ala 180	His	Pro	Pro	Leu	Asn 185	Leu	Ser	Ala	Lys	Thr 190	Glu	Leu
Val	Ala	Trp 195	Ala	Val	Ser	Asn	Trp 200	Lys	Pro	Asp	Ser	Ala 205	Arg	Val	Arg
Tyr	Tyr 210	Gln	Ser	Leu	Gln	Ala 215	His	Leu	Lys	Val	Asp 220	Val	Tyr	Gly	Arg
Ser 225	His	Lys	Pro	Leu	Pro 230	Lys	Gly	Thr	Met	Met 235	Glu	Thr	Leu	Ser	Arg 240
Tyr	Lys	Phe	Tyr	Leu 245	Ala	Phe	Glu	Asn	Ser 250	Leu	His	Pro	Asp	Tyr 255	Ile
Thr	Glu	Lys	Leu 260	Trp	Arg	Asn	Ala	Leu 265	Glu	Ala	Trp	Ala	Val 270	Pr	Val
Val	Leu	Gly 275	Pro	Ser	Arg	Ser	Asn 280	Tyr	Glu	Arg	Phe	Leu 285	Pro	Pro	Asp
Ala 290	Phe	Ile	His	Val	Asp	Asp 295	Phe	Gln	Ser	Pro	Lys 300	Asp	Leu	Ala	Arg
Tyr 305	Leu	Gln	Glu	Leu	Asp 310	Lys	Asp	His	Ala	Arg 315	Tyr	Leu	Ser	Tyr	Phe 320
Arg	Trp	Arg	Glu	Thr 325	Leu	Arg	Pro	Arg	Ser 330	Phe	Ser	Trp	Ala	Leu	Asp 335

Phe Cys Lys Ala Cys Trp Lys Leu Gln Gln Glu S r Arg Tyr Gln Thr
 340 345 350

Val Arg Ser Ile Ala Ala Trp Phe Thr
 355 360

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1500 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCTTCCCTTG TAGACTCTTC TTGGAATGAG AAGTACCGAT TCTGCTGAAG ACCTCGCGCT	60
CTCAGGCTCT GGGAGTTGGA ACCCTGTACC TTCCTTTCCT CTGCTGAGCC CTGCCTCCTT	120
AGGCAGGCCA GAGCTCGACA GAACTCGGTT GCTTTGCTGT TTGCTTTGGA GGGAACACAG	180
CTGACGATGA GGCTGACTTT GAACTCAAGA GATCTGCTTA CCCAGTCTC CTGGAATTAA	240
AGGCCTGTAC TACATTTGCC TGGACCTAAG ATTTTCATGA TCACTATGCT TCAAGATCTC	300
CATGTCAACA AGATCTCCAT GTCAAGATCC AAGTCAGAAA CAAGTCTTCC ATCCTCAAGA	360
TCTGGATCAC AGGAGAAAAT AATGAATGTC AAGGGAAAAG TAATCCTGTT GATGCTGATT	420
GTCTCAACCG TGGTTGTCGT GTTTTGGGAA TATGTCAACA GAATTCCAGA GGTGGTGAG	480
AACAGATGGC AGAAGGACTG GTGGTTCCCA AGCTGGTTTA AAAATGGGAC CCACAGTTAT	540
CAAGAAGACA ACGTAGAAGG ACGGAGAGAA AAGGGTAGAA ATGGAGATCG CATTGAAGAG	600
CCTCAGCTAT GGGACTGGTT CAATCCAAAG AACCGCCCGG ATGTTTTGAC AGTGACCCCG	660
TGGAAGGCGC CGATTGTGTG GGAAGGCACT TATGACACAG CTCTGCTGGA AAAGTACTAC	720
GCCACACAGA AACTCACTGT GGGGCTGACA GTGTTTGCTG TGGGAAAGTA CATTGAGCAT	780
TACTTAGAAG ACTTTCTGGA GTCTGCTGAC ATGTACTTCA TGGTTGGCCA TCGGGTCATA	840
TTTTACGTCA TGATAGACGA CACCTCCCGG ATGCCTGTCTG TGCACCTGAA CCCTCTACAT	900

104250 52429000

TCCTTACAAG TCTTTGAGAT CAGGTCTGAG AAGAGGTGGC AGGATATCAG CATGATGCGC 960
 ATGAAGACCA TTGGGGAGCA CATCCTGGCC CACATCCAGC ACGAGGTCGA CTTCTCTTTC 1020
 TGCATGGACG TGGATCAAGT CTTTCAAGAC AACTTCGGGG TGGAAACTCT GGGCCAGCTG 1080
 GTAGCACAGC TCCAGGCCTG GTGGTACAAG GCCAGTCCCG AGAAGTTCAC CTATGAGAGG 1140
 CGGGAAGTGT CGGCCGCGTA CATTCCATTC GGAGAGGGGG ATTTTACTA CCACGCGGCC 1200
 ATTTTGGAG GAACGCCTAC TCACATTCTC AACCTCACCA GGGAGTGCTT TAAGGGGATC 1260
 CTCCAGGACA AGAAACATGA CATAGAAGCC CAGTGGCATG ATGAGAGCCA CCTCAACAAA 1320
 TACTTCCTTT TCAACAAACC CACTAAAATC CTATCTCCAG AGTATTGCTG GGACTATCAG 1380
 ATAGGCCTGC CTTCAGATAT TAAAGTGTC AAGGTAGCTT GGCAGACAAA AGAGTATAAT 1440
 TTGGTTAGAA ATAATGTCTG ACTTCAAATT GTGATGGAAA CTTGACACTA TTTCTAACCA 1500

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 394 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Ile	Thr	Met	Leu	Gln	Asp	Leu	His	Val	Asn	Lys	Ile	Ser	Met	Ser
1				5					10					15	
Arg	Ser	Lys	Ser	Glu	Thr	Ser	Leu	Pro	Ser	Ser	Arg	Ser	Gly	Ser	Gln
			20					25					30		
Glu	Lys	Ile	Met	Asn	Val	Lys	Gly	Lys	Val	Ile	Leu	Leu	Met	Leu	Ile
		35					40					45			
Val	Ser	Thr	Val	Val	Val	Val	Phe	Trp	Glu	Tyr	Val	Asn	Arg	Il	Pro
	50					55					60				
Glu	Val	Gly	Glu	Asn	Arg	Trp	Gln	Lys	Asp	Trp	Trp	Phe	Pro	Ser	Trp
65					70					75					80
Phe	Lys	Asn	Gly	Thr	His	Ser	Tyr	Gln	Glu	Asp	Asn	Val	Glu	Gly	Arg
				85						90				95	

Arg	Glu	Lys	Gly	Arg	Asn	Gly	Asp	Arg	Ile	Glu	Glu	Pro	Gln	Leu	Trp	
			100					105						110		
Asp	Trp	Phe	Asn	Pro	Lys	Asn	Arg	Pro	Asp	Val	Leu	Thr	Val	Thr	Pro	
		115					120					125				
Trp	Lys	Ala	Pro	Ile	Val	Trp	Glu	Gly	Thr	Tyr	Asp	Thr	Ala	Leu	Leu	
	130					135					140					
Glu	Lys	Tyr	Tyr	Ala	Thr	Gln	Lys	Leu	Thr	Val	Gly	Leu	Thr	Val	Phe	
145					150					155					160	
Ala	Val	Gly	Lys	Tyr	Ile	Glu	His	Tyr	Leu	Glu	Asp	Phe	Leu	Glu	Ser	
			165						170					175		
Ala	Asp	Met	Tyr	Phe	Met	Val	Gly	His	Arg	Val	Ile	Phe	Tyr	Val	Met	
		180						185					190			
Ile	Asp	Asp	Thr	Ser	Arg	Met	Pro	Val	Val	His	Leu	Asn	Pro	Leu	His	
		195					200					205				
Ser	Leu	Gln	Val	Phe	Glu	Ile	Arg	Ser	Glu	Lys	Arg	Trp	Gln	Asp	Ile	
	210					215					220					
Ser	Met	Met	Arg	Met	Lys	Thr	Ile	Gly	Glu	His	Ile	Leu	Ala	His	Ile	
225					230					235					240	
Gln	His	Glu	Val	Asp	Phe	Leu	Phe	Cys	Met	Asp	Val	Asp	Gln	Val	Phe	
				245					250					255		
Gln	Asp	Asn	Phe	Gly	Val	Glu	Thr	Leu	Gly	Gln	Leu	Val	Ala	Gln	Leu	
			260					265					270			
Gln	Ala	Trp	Trp	Tyr	Lys	Ala	Ser	Pro	Glu	Lys	Phe	Thr	Tyr	Glu	Arg	
		275					280					285				
Arg	Glu	Leu	Ser	Ala	Ala	Tyr	Ile	Pro	Phe	Gly	Glu	Gly	Asp	Phe	Tyr	
	290					295					300					
Tyr	His	Ala	Ala	Ile	Phe	Gly	Gly	Thr	Pro	Thr	His	Ile	Leu	Asn	Leu	
305					310					315					320	
Thr	Arg	Glu	Cys	Phe	Lys	Gly	Ile	Leu	Gln	Asp	Lys	Lys	His	Asp	Il	
				325					330					335		
Glu	Ala	Gln	Trp	His	Asp	Glu	Ser	His	Leu	Asn	Lys	Tyr	Phe	Leu	Phe	
			340					345					350			
Asn	Lys	Pro	Thr	Lys	Ile	Leu	Ser	Pro	Glu	Tyr	Cys	Trp	Asp	Tyr	Gln	
		355					360					365				

Ile Gly Leu Pro Ser Asp Ile Lys Ser Val Lys Val Ala Trp Gln Thr
370 375 380

Lys Glu Tyr Asn Leu Val Arg Asn Asn Val
385 390

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8174 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GAATTCCATC GTGGCAAGGG CAGCCTGAAT GGATGATGTA ACCTGGGGTC CTTTCAATGG 60
AGGGCCAGAC TCCTGGGTCT AGGGGATGAG GGAGGGGAGG ATCGGGTTAG CTGGGACCCA 120
GGTGAAAGGG GCTGGGGGCC CACATTCCTG AGTCTCAGAG AGAAGGATCT GGGGTCTCAA 180
GCACCTGAGT CGGAGGGAGG AGGGGTGCTG GGCTCCTGGA AAAACCACCT CTTGGACCAT 240
CTATGCAGAT CACGCAGAAC AAGAGAAATT TCTGCGCCCC ATCTGAATTT CTAAGTTTGG 300
GGGGAGGGCG TGATCTGACA CTGAGGTTCC TTGATCCTCA GCAAGGCGGC AATTGCTGTA 360
TGAAAGAAGC GACCGCATCT GAGACACAAG TATCCTGCCT TGGAAGCCTC TCACCTGGCC 420
GTGGGCCAAC CTCAACCTCA TCTGTCCCTG CTCAGATGCT CAGACCCTGG ACATCCCAGC 480
CTCCTCCTCC CTGATGCAAT CCTGGTGTTT CTTTCACCAG AGAAGCCATC CCAGGCCCAG 540
GCAGGTGCTC CTGAAATAAC CTGGGGGGAG GGGTGGCTGA AAGTCCCTGA CTGGAGTTGG 600
CAGCCAAGCC AGGCCCTGGA GTGGGCACCC AGAGGGAAGA CAGGTTGGCT AATTTCCTGG 660
AGCCCCTAAG GGTGCAAGGG TAGGCCTTCT GTGTCTGAGG GAGGAGGGCT GGGGCTCTGG 720
ACTCCTGGGT CTGAGGGAGG AGGGGTGGGG GGCCTGGACT CCTGGGTCTG AGGGAGGAGG 780
GTCTGGGCCT GTACTCCTGG ATCTGAGGGA GGAGGGGCTG GGGAAGTTGG GCTCCTGGGT 840
CTGAGGGAGG AGGGAGCTTT GGTCTGGACT CCTGGGTCTG AGGGAGTAGG GGCTAGGGAT 900

[illegible]

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2

AGTGCAACCA CACCATTATG ACCATTGGCA CCTTCGGCTT CTGGGCTGCC TACCTGGCTG 5640
GCGGAGACAC TGTCTACCTG GCCAACTTCA CCCTGCCAGA CTCTGAGTTC CTGAAGATCT 5700
TTAAGCCGGA GGCGGCCTTC CTGCCCCGAGT GGGTGGGCAT TAATGCAGAC TTGTCTCCAC 5760
TCTGGACATT GGCTAAGCCT TGAGAGCCAG GGAGACTTTC TGAAGTAGCC TGATCTTTCT 5820
AGAGCCAGCA GTACGTGGCT TCAGAGGCCT GGCATCTTCT GGAGAAGCTT GTGGTGTTC 5880
TGAAGCAAAT GGGTGCCCGT ATCCAGAGTG ATTCTAGTTG GGAGAGTTGG AGAGAAGGGG 5940
GACGTTTCTG GAACTGTCTG AATATTCTAG AACTAGCAA ACATCTTTTC CTGATGGCTG 6000
GCAGGCAGTT CTAGAAGCCA CAGTGCCCAC CTGCTCTTCC CAGCCCATAT CTACAGTACT 6060
TCCAGATGGC TGCCCCCAGG AATGGGGAAC TCTCCCTCTG GTCTACTCTA GAAGAGGGGT 6120
TACTTCTCCC CTGGGTCCTC CAAAGACTGA AGGAGCATAT GATTGCTCCA GAGCAAGCAT 6180
TCACCAAGTC CCCTTCTGTG TTTCTGGAGT GATTCTAGAG GGAGACTTGT TCTAGAGAGG 6240
ACCAGGTTTG ATGCCTGTGA AGAACCCTGC AGGGCCCTTA TGGACAGGAT GGGGTTCTGG 6300
~~AAATCCAGAT AACTAAGGTG AAGAATCTTT TTAGTTTTTT TTTTTTTTTT TTGGAGACAG 6360~~
GGTCTCGCTC TGTTGCCCAG GCTGGAGTGC AGTGGCGTGA TCTTGGCTCA CTGCAACTTC 6420
CGCCTCCTGT GTTCAAGCGA TTCTCCTGTC TCAGCCTCCT GAGTAGATGG GACTACAGGC 6480
ACAGGCCATT ATGCCTGGCT AATTTTTGTA TTTTATAGTAG AGACAGGGTT TCACCATGTT 6540
GGCCGGGATG GTCTCGATCT CCTGACCTTG TCATCCACCT GTCTTGGCCT CCCAAAGTGC 6600
TGGGATTACT GGCATGAGCC ACTGTGCCCA GCCCGGATAT TTTTTTTTAA TTATTTATTT 6660
ATTTATTTAT TTATTGAGAC GGAGTCTTGC TCTGTAGCCC AGGCCAGAGT GCAGTGGCGC 6720
GATCTCAGCT CACTGCAAGC TCTGCCTCCC GGGTTCATGC CATTCTGCCT CAGCCTCCTG 6780
AGTAGCTGGG ACTACAGGCG CCCGCCACCA CGCCCGGCTA ATTTTTTTTG TATTTTTAGT 6840
AGAGACGGGG TTTCATCGTG TTAACCAGGA TGGTCTCGAT CTCCTGACCT CGTGATCTGC 6900
CCACCTCGGC CTCCCACAGT GCTGGGATTA CCGGCGTGAG CCACCATGCC TGGCCCGGAT 6960
AATTTTTTTT AATTTTTGTA GAGACGAGGT CTTGTGATAT TGCCCAGGCT GTTCTTCAAC 7020
TCCTGGGCTC AAGCAGTCCT CCCACCTTGG CCTCCCAGAA TGCTGGGTTT ATAGATGTGA 7080
GCCAGCACAC CGGGCCAAGT GAAGAATCTA ATGAATGTGC AACCTAATTG TAGCATCTAA 7140

104250-524560

TGAATGTTCC ACCATTGCTG GAAAAATTGA GATGGAAAAC AAACCATCTC TAGTTGGCCA	7200
GCGTCTTGCT CTGTTACAG TCTCTGGAAA AGCTGGGGTA GTTGGTGAGC AGAGCGGGAC	7260
TCTGTCCAAC AAGCCCCACA GCCCCTCAA GACTTTTTTTT TGTTTGTTTT GAGCAGACAG	7320
GCTAAAATGT GAACGTGGGG TGAGGGATCA CTGCCAAAAT GGTACAGCTT CTGGAGCAGA	7380
ACTTTCAGG GATCCAGGGA CACTTTTTTTT TAAAGCTCAT AAAGTCCAA GAGCTCCATA	7440
TATTGGGTGT GAGTTCAGGT TGCCTCTCAC AATGAAGGAA GTTGGTCTTT GTCTGCAGGT	7500
GGGCTGCTGA GGGTCTGGGA TCTGTTTTCT GGAAGTGTGC AGGTATAAAC ACACCCTCTG	7560
TGCTTGTGAC AAAGTGGCAG GTACCGTGCT CATTGCTAAC CACTGTCTGT CCCTGAACTC	7620
CCAGAACCAC TACATCTGGC TTTGGGCAGG TCTGAGATAA AACGATCTAA AGGTAGGCAG	7680
ACCCTGGACC CAGCCTCAGA TCCAGGCAGG AGCAGGAGGT CTGGCCAAGG TGGACGGGGT	7740
TGTCGAGATC TCAGGAGCCC CTTGCTGTTT TTTGGAGGGT GAAAGAAGAA ACCTTAAACA	7800
TAGTCAGCTC TGATCACATC CCCTGTCTAC TCATCCAGAC CCCATGCCTG TAGGCTTATC	7860
AGGGAGTTAC AGTTACAATT GTTACAGTAC TGTTCCCAAC TCAGCTGCCA CGGGTGAGAG	7920
AGCAGGAGGT ATGAATTAAA AGTCTACAGC ACTAACCCGT GTCTCTGTAG CTTTTTTGGA	7980
GCCAGAGCCA CTGTGTATGT GTGTGTGGGT TTGTGTGTGT GTGTGTGTGT GTGTGTGTGT	8040
AAGAGAGTGG AGGAAAAGGT GGGGTACTTC TGAAGACTTT TATTTTTTTTT TAATTAATTT	8100
ATTTTTTTTC AGAGATCGAG TCTTGCTCTG TGGCCCAGGC TGGAGTGCAG TAGTGTGATC	8160
TCGGCCCACT GCAA	8170

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 365 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met	Trp	Leu	Arg	Ser	His	Arg	Gln	Leu	Cys	Leu	Ala	Phe	Leu	Leu	Val
1				5					10				15		

[illegible]

Thr	Pro	Trp	Lys	Asp	Phe	Ala	Leu	Leu	Thr	Gln	Cys	Asn	His	Thr	Ile
290						295					300				
Met	Thr	Ile	Gly	Thr	Phe	Gly	Phe	Trp	Ala	Ala	Tyr	Leu	Ala	Gly	Gly
305					310				315					320	
Asp	Thr	Val	Tyr	Leu	Ala	Asn	Phe	Thr	Leu	Pro	Asp	Ser	Glu	Phe	Leu
				325					330					335	
Lys	Ile	Phe	Lys	Pro	Glu	Ala	Ala	Phe	Leu	Pro	Glu	Trp	Val	Gly	Ile
			340					345					350		
Asn	Ala	Asp	Leu	Ser	Pro	Leu	Trp	Thr	Leu	Ala	Lys	Pro			
		355					360					365			

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3647 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTGCAGAGAG CGCCACCCGG AAGCCACTTT TATAGAAGCT TTTACACACA ATGCTTGATT	60
TTTTTTTTTT TTTTCCGAGA CGGAGTCTCG CTTTGTCGCC CAGGCTGGAG TGCAGTGGCG	120
CGATCTGGGC TCACTGCAAG CTCCGCCTCC TGGGTTGACG CCATTCTCCT GCCTCAGCTT	180
CCCGAGTAGC TGGGACTACA GGCGCCCGCC ACCAAGCCTG GCTAATTTTT TTTTATTTTT	240
AGTGGAGACA GAGTTTCACC GTGTTAGCCA GGATGGTCTC GATCTCCTGA CCTCGGGATC	300
CGCCCGCCTC GGCCTCCCAA AGTGCTGGGA GTATAGGCGT GAGCCACCGC GCCTGGCCTA	360
TACTTGATTT TTAATGAAAA CATTCTTAA TTCATATGGC TAACGCAAAT TTATTTTCTG	420
TAGGCATAAC ATCAAAAACA CCTGGCAGGA CTGCCCCATT CCCAGCACTG TCTAGTTCTC	480
CCCTAGTATC AGTGGGACTC CACTGATGCA CAGCTGTGAT CTACTAAAAC TTCTCTCAAA	540
ACTTTCTCCT CTCCTTAGGT CAGCAGCCCC GCCCCTGATC TATTTGGAAA TCCCCTGAAT	600

0063475-0540

AAAAGTTGAA	TATCATAAAC	CAAAGCGAAC	ACCCAGAAAT	TCAAATTCAA	CCCGTAGGTA	66
AAAAATTTCT	CAAGTGACTG	TAGACGTAGA	TGTCTCCAGT	GTCGCCTAAT	AAGGTAGAAG	72
AGGCCAGTGC	GATACTGTCT	TTACACCCTT	AACTTGGGTG	CTAGAATATT	TATCTTCGTC	78
ATCATTTTAT	CATCCAAACT	ATTTTGCATA	ACTTTCATGG	GTGCAGAAAA	TGTTTTTTTAA	84
GTGCTTGGTA	AAATTAATAG	TGATATTCAT	TCATTCATCT	CACTGAACAG	GCAATAAATT	90
CCTTGACGAC	AAGGGCCTTG	GGGGGGGCCA	CATCTTCATC	TTTGGTTTAT	GAGTCCTGTG	96
CGTCTTGGTA	CAAGCAATAC	TACTATGAGC	CGGCAAGTCA	GACTTATTTG	GTAGGGGACC	102
AAAGGAAAGA	ACATGTTTTG	ATTGCTAAGA	AAACATTTTG	TTCTCTATCC	TTTACTGGGC	108
TGGCAGGCAA	AGGAAATGTT	CTTATGAGCA	CTCACATTGA	AAACTTAAGT	TCTTCACCAA	114
ATGCAGAGAC	TCTGAAGGCC	ACGCCGCTGC	GGGCTGCCTC	CACAATTCGA	CCGTCTCGGC	120
GGGCCACGAG	ATCCTGGCCA	CGGATGCGGT	GGCCGCGCCT	CTGCTCGCAC	GTTCCCCCGG	126
CCTCTGGACT	CCCTCCCTCC	CTCAATCCCT	CCCTCCGGCG	GGCGTCGCTG	GCGGGTGGCT	132
AGGCCCAACG	GCAGGAAGCC	GACGCTATCC	TCCGTTCCGC	GGCGCCGGGT	CCGCCTTCCG	138
TCTGTTCTAG	GGCCTGCTCC	TGCGCGGCAG	CTGCTTTAGA	AGGTCTCGAG	CCTCCTGTAC	144
CTTCCCAGGG	ATGAACCGGG	CCTTCCCTCT	GGAAGGCGAG	GGTTCGGGCC	ACAGTGAGCG	150
AGGGCCAGGG	CGGTGGGCGC	GCGCAGAGGG	AAACCGGATC	AGTTGAGAGA	GAATCAAGAG	156
TAGCGGATGA	GGCGCTTG TG	GGGCGCGGCC	CGGAAGCCCT	CGGGCGCGGG	CTGGGAGAAG	162
GAGTGGGCGG	AGGCGCCGCA	GGAGGCTCCC	GGGGCCTGGT	CGGGCCGGCT	GGGCCCCGGG	168
CGCAGTGGAA	GAAAGGGACG	GGCGGTGCCC	GGTTGGGCGT	CCTGGCCAGC	TCACCTTGCC	174
CTGGCGGGCTC	GCCCCGCCCCG	GCACTTGGGA	GGAGCAGGGC	AGGGCCCCGCG	GCCTTTGCAT	180
TCTGGGACCG	CCCCCTTCCA	TTCCCGGGCC	AGCGGCGAGC	GGCAGCGACG	GCTGGAGCCG	186
CAGCTACAGC	ATGAGAGCCG	GTGCCGCTCC	TCCACGCCTG	CGGACGCGTG	GCGAGCGGAG	192
GCAGCGCTGC	CTGTTGCGCG	CATGGGGGCA	CCGTGGGGCT	CGCCGACGGC	GGCGGCGGGC	198
GGGCGGCGCG	GGTGGCGCCG	AGGCCGGGGG	CTGCCATGGA	CCGTCTGTGT	GCTGGCGGCC	204
GCCGGCTTGA	CGTGACGGC	GCTGATCACC	TACGCTTGCT	GGGGGCAGCT	GCCGCCGCTG	210
CCCTGGGCGT	CGCCAACCCC	GTCGCGACCG	GTGGGCGTGC	TGCTGTGGTG	GGAGCCCTTC	216

GGGGGGCGCG	ATAGCGCCCC	GAGGCCGCCC	CCTGACTGCC	CGCTGCGCTT	CAACATCAGC	2220
GGCTGCCGCC	TGCTCACCGA	CCGCGCGTCC	TACGGAGAGG	CTCAGGCCGT	GCTTTTCCAC	2280
CACCGCGACC	TCGTGAAGGG	GCCCCCGAC	TGGCCCCCGC	CCTGGGGCAT	CCAGGCGCAC	2340
ACTGCCGAGG	AGGTGGATCT	GCGCGTGTG	GACTACGAGG	AGGCAGCGGC	GGCGGCAGAA	2400
GCCCTGGCGA	CCTCCAGCCC	CAGGCCCCCG	GGCCAGCGCT	GGGTTTGGAT	GAACTTCGAG	2460
TCGCCCTCGC	ACTCCCCGGG	GCTGCGAAGC	CTGGCAAGTA	ACCTCTTCAA	CTGGACGCTC	2520
TCCTACCGGG	CGGACTCGGA	CGTCTTTGTG	CCTTATGGCT	ACCTCTACCC	CAGAAGCCAC	2580
CCCGGCGACC	CGCCCTCAGG	CCTGGCCCCG	CCACTGTCCA	GGAAACAGGG	GCTGGTGGCA	2640
TGGGTGGTGA	GCCACTGGGA	CGACCGCCAG	GCCCGGGTCC	GCTACTACCA	CCAACTGAGC	2700
CAACATGTGA	CCGTGGACGT	GTTCGGCCGG	GGCGGGCCGG	GGCAGCCGGT	GCCCGAAATT	2760
GGGCTCCTGC	ACACAGTGGC	CCGCTACAAG	TTCTACCTGG	CTTTCGAGAA	CTCGCAGCAC	2820
CTGGATTATA	TCACCGAGAA	GCTCTGGCGC	AACGCGTTGC	TCGCTGGGGC	GGTGCCGGTG	2880
GTGCTGGGCC	CAGACCGTGC	CAACTACGAG	GCGTTTGTGC	CCCGCGGCGC	CTTCATCCAC	2940
GTGGACGACT	TCCCAAGTGC	CTCCTCCCTG	GCCTCGTACC	TGCTTTTCTT	CGACCGCAAC	3000
CCCGCGGTCT	ATCGCCGCTA	CTTCCACTGG	CGCCGGAGCT	ACGCTGTCCA	CATCACCTCC	3060
TTCTGGGACG	AGCCTTGGTG	CCGGGTGTGC	CAGGCTGTAC	AGAGGGCTGG	GGACCGGCCC	3120
AAGAGCATA	GGAACCTGGC	CAGCTGGTTC	GAGCGGTGAA	GCCGCGCTCC	CCTGGAAGCG	3180
ACCCAGGGGA	GCCCAAGTTG	TCAGCTTTTT	GATCCTCTAC	TGTGCATCTC	CTTGACTGCC	3240
GCATCATGGG	AGTAAGTTCT	TCAAACACCC	ATTTTGTGCTC	TATGGGAAAA	AAACGATTTA	3300
CCAATTAATA	TTACTCAGCA	CAGAGATGGG	GGCCCGGTTT	CCATATTTTT	TGCACAGCTA	3360
GCAATTGGGC	TCCCTTTGCT	GCTGATGGGC	ATCATTGTTT	AGGGGTGAAG	GAGGGGGTTC	3420
TTCCTCACCT	TGTAACCAGT	GCAGAAATGA	AATAGCTTAG	CGGCAAGAAG	CCGTTGAGGC	3480
GGTTTCCTGA	ATTTCCCAT	CTGCCACAGG	CCATATTTGT	GGCCCGTGCA	GCTTCCAAAT	3540
CTCATACACA	ACTGTTCCCG	ATTCACGTTT	TTCTGGACCA	AGGTGAAGCA	AATTTGTGGT	3600
TGTAGAAGGA	GCCTTGTTGG	TGGAGAGTGG	AAGGACTGTG	GCTGCAG		3660

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 405 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met	Gly	Ala	Pro	Trp	Gly	Ser	Pro	Thr	Ala	Ala	Ala	Gly	Gly	Arg	Arg	1	5	10	15
Gly	Trp	Arg	Arg	Gly	Arg	Gly	Leu	Pro	Trp	Thr	Val	Cys	Val	Leu	Ala	20	25	30	
Ala	Ala	Gly	Leu	Thr	Cys	Thr	Ala	Leu	Ile	Thr	Tyr	Ala	Cys	Trp	Gly	35	40	45	
Gln	Leu	Pro	Pro	Leu	Pro	Trp	Ala	Ser	Pro	Thr	Pro	Ser	Arg	Pro	Val	50	55	60	
Gly	Val	Leu	Leu	Trp	Trp	Glu	Pro	Phe	Gly	Gly	Arg	Asp	Ser	Ala	Pr	65	70	75	80
Arg	Pro	Pro	Pro	Asp	Cys	Pro	Leu	Arg	Phe	Asn	Ile	Ser	Gly	Cys	Arg	85	90	95	
Leu	Leu	Thr	Asp	Arg	Ala	Ser	Tyr	Gly	Glu	Ala	Gln	Ala	Val	Leu	Phe	100	105	110	
His	His	Arg	Asp	Leu	Val	Lys	Gly	Pro	Pro	Asp	Trp	Pro	Pro	Pr	Trp	115	120	125	
Gly	Ile	Gln	Ala	His	Thr	Ala	Glu	Glu	Val	Asp	Leu	Arg	Val	Leu	Asp	130	135	140	
Tyr	Glu	Glu	Ala	Ala	Ala	Ala	Ala	Glu	Ala	Leu	Ala	Thr	Ser	Ser	Pro	145	150	155	160
Arg	Pro	Pro	Gly	Gln	Arg	Trp	Val	Trp	Met	Asn	Phe	Glu	Ser	Pr	Ser	165	170	175	
His	Ser	Pro	Gly	Leu	Arg	Ser	Leu	Ala	Ser	Asn	Leu	Phe	Asn	Trp	Thr	180	185	190	
Leu	Ser	Tyr	Arg	Ala	Asp	Ser	Asp	Val	Phe	Val	Pro	Tyr	Gly	Tyr	Leu	195	200	205	

006347-0540

Tyr Pro Arg Ser His Pro Gly Asp Pro Pro Ser Gly Leu Ala Pro Pro
 210 215 220
 Leu Ser Arg Lys Gln Gly Leu Val Ala Trp Val Val Ser His Trp Asp
 225 230 235 240
 Glu Arg Gln Ala Arg Val Arg Tyr Tyr His Gln Leu Ser Gln His Val
 245 250 255
 Thr Val Asp Val Phe Gly Arg Gly Gly Pro Gly Gln Pro Val Pro Glu
 260 265 270
 Ile Gly Leu Leu His Thr Val Ala Arg Tyr Lys Phe Tyr Leu Ala Phe
 275 280 285
 Glu Asn Ser Gln His Leu Asp Tyr Ile Thr Glu Lys Leu Trp Arg Asn
 290 295 300
 Ala Leu Leu Ala Gly Ala Val Pro Val Val Leu Gly Pro Asp Arg Ala
 305 310 315 320
 Asn Tyr Glu Arg Phe Val Pro Arg Gly Ala Phe Ile His Val Asp Asp
 325 330 335
 Phe Pro Ser Ala Ser Ser Leu Ala Ser Tyr Leu Leu Phe Leu Asp Arg
 340 345 350
 Asn Pro Ala Val Tyr Arg Arg Tyr Phe His Trp Arg Arg Ser Tyr Ala
 355 360 365
 Val His Ile Thr Ser Phe Trp Asp Glu Pro Trp Cys Arg Val Cys Gln
 370 375 380
 Ala Val Gln Arg Ala Gly Asp Arg Pro Lys Ser Ile Arg Asn Leu Ala
 385 390 395 400
 Ser Trp Phe Glu Arg
 405

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1488 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGGGGGCAC CGTGGGGCTC GCCGACGGCG GCGGCGGGCG GCGGGCGCGG GTGGCGCCGA 60
GGCCCGGGGC TGCCATGGAC CGTCTGTGTG CTGGCGGGCG CCGGCTTGAC GTGTACGGCG 120
CTGATCACCT ACGCTTGCTG GGGGCAGCTG CCGCCGCTGC CCTGGGCGTC GCCAACCCCG 180
TCGCGACCGG TGGGCGTGCT GCTGTGGTGG GAGCCCTTCG GGGGGCGCGA TAGCGCCCCG 240
AGGCCGCCCC CTGACTGCTG CTGGGGGCAG CTGCCGCCGC TGCCCTGGGC GTCGCCAACC 300
CCGTCGCGAC CGGTGGGCGT GCTGCTGTGG TGGGAGCCCT TCGGGGGGCG CGATAGCGCC 360
CCGAGGCCGC CCCCTGACTG CCCGCTGCGC TTCAACATCA GCGGCTGCCG CCTGCTCACC 420
GACCGCGCGT CCTACGGAGA GGCTCAGGCC GTGCTTTTCC ACCACCGCGA CCTCGTGAAG 480
GGGCCCCCGG ACTGGCCCCC GCCCTGGGGC ATCCAGGCGC AACTTGCCGA GCCGCTGCGC 540
TTCAACATCA GCGGCTGCCG CCTGCTCACC GACCGCGCGT CCTACGGAGA GGCTCAGGCC 600
GTGCTTTTCC ACCACCGCGA CCTCGTGAAG GGGCCCCCGG ACTGGCCCCC GCCCTGGGGC 660
ATCCAGGCGC AACTTGCCGA GGAGGTGGAT CTGCGCGTGT TGGACTACGA GGAGGCAGCG 720
GCGGCGGCAG AAGCCCTGGC GACCTCCAGC CCCAGGCCCC CGGGCCAGCG CTGGGTTTGG 780
ATGAACTTCG AGTCGCCCTC GCACTCCCCG GGGCTGCGAA GCCTGGCAAG TAACCTCTTC 840
AACTGGACGC TCTCCTACCG GGCGGACTCG GACGTCTTTG TGCCTTATGG CTACCTCTAC 900
CCCAGAAGCC ACCCCGGCGA CCCGCCCTCA GGCCTGGCCC CGCCACTGTC CAGGAAACAG 960
GGGCTGGTGG CATGGGTGGT GAGCCACTGG GACGAGCGCC AGGCCCGGGT CCGCTACTAC 1020
CACCAACTGA GCCAACATGT GACCGTGGAC GTGTTGCGCC GGGGCGGGCC GGGGCAGCCG 1080
GTGCCCCGAA TTGGGCTCCT GCACACAGTG GCCCGCTACA AGTTCTACCT GGCTTTCGAG 1140
AACTCGCAGC ACCTGGATTA TATCACCAGG AAGCTCTGGC GCAACGCGTT GCTCGCTGGG 1200
GCGGTGCCGG TGGTGCTGGG CCCAGACCGT GCCAACTACG AGCGCTTTGT GCCCCGCGGC 1260
GCCTTCATCC ACGTGGACGA CTTCCCAAGT GCCTCCTCCC TGGCCTCGTA CCTGCTTTTC 1320
CTCGACCGCA ACCCCGCGGT CTATCGCCGC TACTTCCACT GGCGCCGGAG CTACGCTGTC 1380
CACATCACCT CCTTCTGGGA CGAGCCTTGG TGCCGGGTGT GCCAGGCTGT ACAGAGGGCT 1440
GGGGACCGGC CCAAGAGCAT ACGGAACTTG GCCAGCTGGT TCGAGCGG 1480

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1316 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTTATGACAA GCTGTGTCAT AAATTATAAC AGCTTCTCTC AGGACACTGT GGCCAGGAAG 60
TGGGTGATCT TCCTTAATGA CCCTCACTCC TCTCTCCTCT CTTCCTCAGCT ACTCTGACCC 120
ATGGATCCCC TGGGCCCAGC CAAGCCACAG TGGCTGTGGC GCCGCTGTCT GGCCGGGCTG 180
CTGTTTCAGC TGCTGGTGGC TGTGTGTTTTC TTCTCCTACC TGCCTGTGTC CCGAGACGAT 240
GCCACTGGAT CCCCTAGGCC AGGGCTTATG GCAGTGGAAC CTGTCACCGG GGCTCCCAAT 300
GGGTCCCGCT GCCAGGACAG CATGGCGACC CCTGCCCACC CCACCCTACT GATCCTGCTG 360
TGGACGTGGC CTTTAAACAC ACCCGTGGCT CTGCCCCGCT GCTCAGAGAT GGTGCCCCGGC 420
GCGGCCGACT GCAACATCAC TGCCGACTCC AGTGTGTACC CACAGGCAGA CGCGGTCATC 480
GTGCACCACT GGGATATCAT GTACAACCCC AGTGCCAACC TCCCGCCCCC CACCAGGCCG 540
CAGGGGCAGC GCTGGATCTG GTTCAGCATG GAGTCCCCCA GCAACTGCCG GCACCTGGAA 600
GCCCTGGACG GATACTTCAA TCTCACCATG TCCTACCGCA GCGACTCCGA CATCTTCACG 660
CCCTACGGCT GGCTGGAGCC GTGGTCCGGC CAGCCTGCCC ACCCACCCTG CAACCTCTCG 720
GCCAAGACCG AGCTGGTGGC CTGGGCGGTG TCCAAGTGA AGCCGGACTC GGCCAGGGTG 780
CGCTACTACC AGAGCCTGCA GGCTCATCTC AAGGTGGACG TGTACGGACG CTCCCACAAG 840
CCCCTGCCCCA AGGGGACCAT GATGGAGACG CTGTCCCGGT ACAAGTTCTA TCTGGCCTTC 900
GAGAACTCCT TGCACCCCGA CTACATCACC GAGAAGCTGT GGAGGAACGC CCTGGAGGCC 960
TGGGCCGTGC CCGTGGTGCT GGGCCCCAGC AGAAGCAACT ACGAGAGGTT CCTGCCGCCC 1020
GACGCCTTCA TCCACGTGGA TGAATTCCAG AGCCCCAAGG ACCTGGCCCCG GTACCTGCAG 1080
GAGCTGGACA AGGACCACGC CCGCTACCTG AGCTACTTTC GCTGGCGGGA GACGCTGCGG 1140

CCTCGCTCCT TCAGCTGGGC ACTGGCTTTC TGCAAGGCCT GCTGGAAGCT GCAGCAGGAA 1200
TCCAGGTACC AGACGGTGCG CAGCATAGCG GCTTG GTTCA CCTGAGAGGC CGGCATGGGG 1260
CCTGGGCTGC CAGGGACCTC ACTTTCCAG GGCCTCACCT ACCTAGGGTC TCTAGA 1316

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met	Asp	Pro	Leu	Gly	Pro	Ala	Lys	Pro	Gln	Trp	Leu	Trp	Arg	Arg	Cys
1				5					10					15	
Leu	Ala	Gly	Leu	Leu	Phe	Gln	Leu	Leu	Val	Ala	Val	Cys	Phe	Phe	Ser
			20					25					30		
Tyr	Leu	Arg	Val	Ser	Arg	Asp	Asp	Ala	Thr	Gly	Ser	Pro	Arg	Pro	Gly
		35					40					45			
Leu	Met	Ala	Val	Glu	Pro	Val	Thr	Gly	Ala	Pro	Asn	Gly	Ser	Arg	Cys
	50					55					60				
Gln	Asp	Ser	Met	Ala	Thr	Pro	Ala	His	Pro	Thr	Leu	Leu	Ile	Leu	Leu
65					70					75					80
Trp	Thr	Trp	Pro	Phe	Asn	Thr	Pro	Val	Ala	Leu	Pro	Arg	Cys	Ser	Glu
				85					90					95	
Met	Val	Pro	Gly	Ala	Ala	Asp	Cys	Asn	Ile	Thr	Ala	Asp	Ser	Ser	Val
			100					105					110		
Tyr	Pro	Gln	Ala	Asp	Ala	Val	Ile	Val	His	His	Trp	Asp	Ile	Met	Tyr
		115					120					125			
Asn	Pro	Ser	Ala	Asn	Leu	Pro	Pro	Pro	Thr	Arg	Pro	Gln	Gly	Gln	Arg
	130					135						140			
Trp	Ile	Trp	Phe	Ser	Met	Glu	Ser	Pro	Ser	Asn	Cys	Arg	His	Leu	Glu
145					150					155					160
Ala	Leu	Asp	Gly	Tyr	Phe	Asn	Leu	Thr	Met	Ser	Tyr	Arg	Ser	Asp	Ser
				165					170					175	

Asp Ile Phe Thr Pro Tyr Gly Trp Leu Glu Pro Trp Ser Gly Gln Pro
180 185 190

Ala His Pro Pro Leu Asn Leu Ser Ala Lys Thr Glu Leu Val Ala Trp
195 200 205

Ala Val Ser Asn Trp Lys Pro Asp Ser Ala Arg Val Arg Tyr Tyr Gln
210 215 220

Ser Leu Gln Ala His Leu Lys Val Asp Val Tyr Gly Arg Ser His Lys
225 230 235 240

Pro Leu Pro Lys Gly Thr Met Met Glu Thr Leu Ser Arg Tyr Lys Phe
245 250 255

Tyr Leu Ala Phe Gln Asn Ser Leu His Pro Asp Tyr Ile Thr Glu Lys
260 265 270

Leu Trp Arg Asn Ala Leu Glu Ala Trp Ala Val Pro Val Val Leu Gly
275 280 285

Pro Ser Arg Ser Asn Tyr Glu Arg Phe Leu Pro Pro Asp Ala Phe Ile
290 295 300

His Val Asp Asp Phe Gln Ser Pro Lys Asp Leu Ala Arg Tyr Leu Gln
305 310 315 320

Glu Leu Asp Lys Asp His Ala Arg Tyr Leu Ser Tyr Phe Arg Trp Arg
325 330 335

Glu Thr Leu Arg Pro Arg Ser Phe Ser Trp Ala Leu Ala Phe Cys Lys
340 345 350

Ala Cys Trp Lys Leu Gln Gln Glu Ser Arg Tyr Gln Thr Val Arg Ser
355 360 365

Ile Ala Ala Trp Phe Thr
370

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1086 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATGGATCCCC TGGGTGCAGC CAAGCCACAA TGGCCATGGC GCCGCTGTCT GGCCGCACTG	60
CTATTTTCAGC TGCTGGTGGC TGTGTGTTTC TTCTCCTACC TGCCTGTGTC CCGAGACGAT	120
GCCACTGGAT CCCCTAGGGC TCCCAGTGGG TCCTCCCGAC AGGACACCAC TCCCACCCGC	180
CCCACCCTCC TGATCCTGCT ATGGACATGG CCTTTCCACA TCCCTGTGGC TCTGTCCCGC	240
TGTTTCAGAGA TGGTGCCCGG CACAGCCGAC TGCCACATCA CTGCCGACCG CAAGGTGTAC	300
CCACAGGCAG ACACGGTCAT CGTGCACCAC TGGGATATCA TGTCCAACCC TAAGTCACGC	360
CTCCCACCTT CCCCAGAGGC GCAGGGGCAG CGCTGGATCT GGTTCAACTT GGAGCCACCC	420
CCTAACTGCC AGCACCTGGA AGCCCTGGAC AGATACTTCA ATCTCACCAT GTCCTACCGC	480
AGCGACTCCG ACATCTTCAC GCCCTACGGC TGGCTGGAGC CGTGGTCCGG CCAGCCTGCC	540
CACCCACCGC TCAACCTCTC GGCCAAGACC GAGCTGGTGG CCTGGGCGGT GTCCAAGTGG	600
AAGCCGGACT CAGCCAGGGT GCGCTACTAC CAGAGCCTGC AGGCTCATCT CAAGGTGGAC	660
GTGTACGGAC GCTCCCACAA GCCCCTGCCC AAGGGGACCA TGATGGAGAC GCTGTCCCGG	720
TACAAGTTCT ACCTGGCCTT CGAGAACTCC TTGCACCCCG ACTACATCAC CGAGAAGCTG	780
TGGAGGAACG CCCTGGAGGC CTGGGCCGTG CCCGTGGTGC TGGGCCCCAG CAGAAGCAAC	840
TACGAGAGGT TCCTGCCACC CGACGCCTTC ATCCACGTGG ACGACTTCCA GAGCCCCAAG	900
GACCTGGCCC GGTACCTGCA GGAGCTGGAC AAGGACCACG CCCGCTACCT GAGCTACTTT	960
CGCTGGCGGG AGACGCTGCG GCCTCGCTCC TTCAGCTGGG CACTGGATTT CTGCAAGGCC	1020
TGCTGGAAAC TGCAGCAGGA ATCCAGGTAC CAGACGGTGC GCAGCATAGC GGCTTGGTTC	1080
ACCTGA	1080

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1654 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TTTTCTCATC TGTGAAACAG GAATAATAAC AGCTCTTCTC AGGACTCATG GCCTGGAGCT	60
TTGGTAAGCA GGAGATTGTC ATCAATGACC CTCACTCCTC TCTCCCCACT TCCCAGAGAC	120
TCTGACCCAT GGATCCCCTG GGCCCGGCCA AGCCACAGTG GTCGTGGCGC TGCTGTCTGA	180
CCACGCTGCT GTTTCAGCTG CTGATGGCTG TGTGTTTCTT CTCCTATCTG CGTGTGTCTC	240
AAGACGATCC CACTGTGTAC CCTAATGGGT CCCGCTTCCC AGACAGCACA GGGACCCCCG	300
CCCACTCCAT CCCCCTGATC CTGCTGTGGA CGTGGCCTTT TAACAAACCC ATAGCTCTGC	360
CCCGCTGCTC AGAGATGGTG CCTGGCACGG CTGACTGCAA CATCACTGCC GACCGCAAGG	420
TGTATCCACA GGCAGACGCG GTCATCGTGC ACCACCGAGA GGTCATGTAC AACCCAGTG	480
CCCAGCTCCC ACGCTCCCCG AGGCGGCAGG GGCAGCGATG GATCTGGTTC AGCATGGAGT	540
CCCCAAGCCA CTGCTGGCAG CTGAAAGCCA TGGACGGATA CTTCAATCTC ACCATGTCCT	600
AACCGCAGCGA CTCCGACATC TTCACGCCCT ACGGCTGGCT GGAGCCGTGG TCCGGCCAGC	660
CTGCCCACCC ACCGCTCAAC CTCTCGGCCA AGACCGAGCT GGTGGCCTGG GCAGTGTCCA	720
ACTGGGGGCC AAACCTCCGCC AGGGTGCGCT ACTACCAGAG CCTGCAGGCC CATCTCAAGG	780
TGGACGTGTA CGGACGCTCC CACAAGCCCC TGCCCCAGGG AACCATGATG GAGACGCTGT	840
CCCGGTACAA GTTCTATCTG GCCTTCGAGA ACTCCTTGCA CCCCAGCTAC ATCACCGAGA	900
AGCTGTGGAG GAACGCCCTG GAGGCCTGGG CCGTGCCCGT GGTGCTGGGC CCCAGCAGAA	960
GCAACTACGA GAGGTTCTTG CCACCCGACG CTTTCATCCA CGTGGACGAC TTCCAGAGCC	1020
CCAAGGACCT GGCCCGGTAC CTGCAGGAGC TGGACAAGGA CCACGCCCGC TACCTGAGCT	1080
ACTTTCGCTG GCGGGAGACG CTGCGGCCTC GCTCCTTCAG CTGGGCACTC GCTTTCTGCA	1140
AGGCCTGCTG GAAACTGCAG GAGGAATCCA GGTACCAGAC ACGCGGCATA GCGGCTTGGT	1200
TCACCTGAGA GGCTGGTGTG GGGCCTGGGC TGCCAGGAAC CTCATTTTCC TGGGGCCTCA	1260
CCTGAGTGGG GGCCTCATCT ACCTAAGGAC TCGTTTGCCT GAAGCTTCAC CTGCCTGAGG	1320
ACTCACCTGC CTGGGACGGT CACCTGTTGC AGCTTCACCT GCCTGGGGAT TCACCTACCT	1380
GGGTCCTCAC TTTCCTGGGG CCTCACCTGC TGGAGTCTTC GGTGGCCAGG TATGTCCCTT	1440
ACCTGGGATT TCACATGCTG GCTTCCAGGA GCGTCCCCTG CGGAAGCCTG GCCTGCTGGG	1500

GATGTCTCCT GGGGACTTTG CCTACTGGGG ACCTCGGCTG TTGGGGACTT TACCTGCTGG 1560
 GACCTGCTCC CAGAGACCTT CCACACTGAA TCTCACCTGC TAGGAGCCTC ACCTGCTGGG 1620
 GACCTCACCC TGGAGGCACT GGGCCCTGGG AACT 1654

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 359 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met	Asp	Pro	Leu	Gly	Pro	Ala	Lys	Pro	Gln	Trp	Ser	Trp	Arg	Cys	Cys	1	5	10	15
Leu	Thr	Thr	Leu	Leu	Phe	Gln	Leu	Leu	Met	Ala	Val	Cys	Phe	Phe	Ser	20	25	30	
Tyr	Leu	Arg	Val	Ser	Gln	Asp	Asp	Pro	Thr	Val	Tyr	Pro	Asn	Gly	Ser	35	40	45	
Arg	Phe	Pro	Asp	Ser	Thr	Gly	Thr	Pro	Ala	His	Ser	Ile	Pro	Leu	Ile	50	55	60	
Leu	Leu	Trp	Thr	Trp	Pro	Phe	Asn	Lys	Pro	Ile	Ala	Leu	Pro	Arg	Cys	65	70	75	
Ser	Glu	Met	Val	Pro	Gly	Thr	Ala	Asp	Cys	Asn	Ile	Thr	Ala	Asp	Arg	85	90	95	
Lys	Val	Tyr	Pro	Gln	Ala	Asp	Ala	Val	Ile	Val	His	His	Arg	Glu	Val	100	105	110	
Met	Tyr	Asn	Pro	Ser	Ala	Gln	Leu	Pro	Arg	Ser	Pro	Arg	Arg	Gln	Gly	115	120	125	
Gln	Arg	Trp	Ile	Trp	Phe	Ser	Met	Glu	Ser	Pro	Ser	His	Cys	Trp	Gln	130	135	140	
Leu	Lys	Ala	Met	Asp	Gly	Tyr	Phe	Asn	Leu	Thr	Met	Ser	Tyr	Arg	Ser	145	150	155	
Asp	Ser	Asp	Ile	Phe	Thr	Pro	Tyr	Gly	Trp	Leu	Glu	Pro	Trp	Ser	Gly	165	170	175	

